

*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Solos  
Ministério da Agricultura, Pecuária e Abastecimento  
Secretaria de Desenvolvimento Agropecuário e Cooperativismo  
Departamento de Propriedade Intelectual e Tecnologia da Agropecuária*

# **GEOESTATÍSTICA APLICADA NA AGRICULTURA DE PRECISÃO UTILIZANDO O VESPER**

*Ronaldo Pereira de Oliveira  
Célia Regina Grego  
Ziany Neiva Brandão  
Editores Técnicos*

**Embrapa**  
*Brasília, DF*  
2015

## Capítulo 2

---

# Pré-processamento dos Dados

Célia Regina Grego

Ronaldo Pereira de Oliveira

### 2.1 Análise Exploratória

Como nas técnicas estatísticas, a geoestatística baseia-se em um conceito probabilístico. Por tanto, para o uso apropriado da geoestatística é necessário um embasamento teórico prévio sobre os fundamentos estatísticos, possibilitando assim que se faça uma análise estatística elementar das variáveis quantitativas a serem submetidas à análise variográfica e interpolação. Uma análise de consistência sobre os resultados da estatística descritiva é fundamental para avaliar se as condições exigidas pelo formalismo matemático e algoritmos da geoestatística são satisfeitas pelos dados de entrada. Estes procedimentos básicos indicam medidas que sintetizam a informação quantitativa contida no acervo de dados brutos, permitindo identificar dados discrepantes (i.e.: “*outliers*”) e possíveis modelos a serem utilizados numa inferência estatística. As medidas resumo caracterizam a tendência central e a dispersão numérica dos valores observados (e.g.: máximo, mínimo, média, mediana, desvio padrão, coeficiente de variação) e fazem parte do conjunto de funções básicas disponíveis numa planilha eletrônica, como demonstrado neste capítulo. Segundo Grego et al. (2014), a etapa preliminar de análise exploratória dos dados é de suma importância e tem o objetivo de identificar previamente a normalidade de distribuição de frequência das medidas e a variação nos dados.

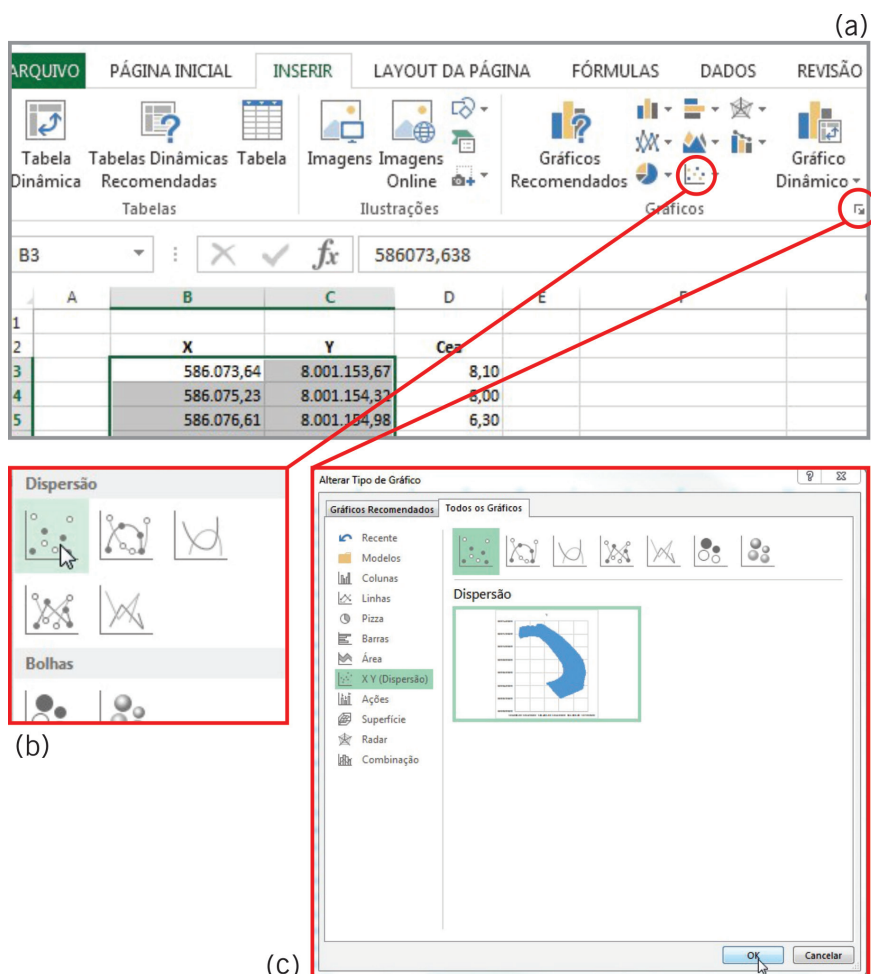
### 2.2 Localização das Observações

O objetivo deste procedimento é ter os dados plotados numa representação gráfica segundo suas coordenadas X e Y (i.e.: coorde-

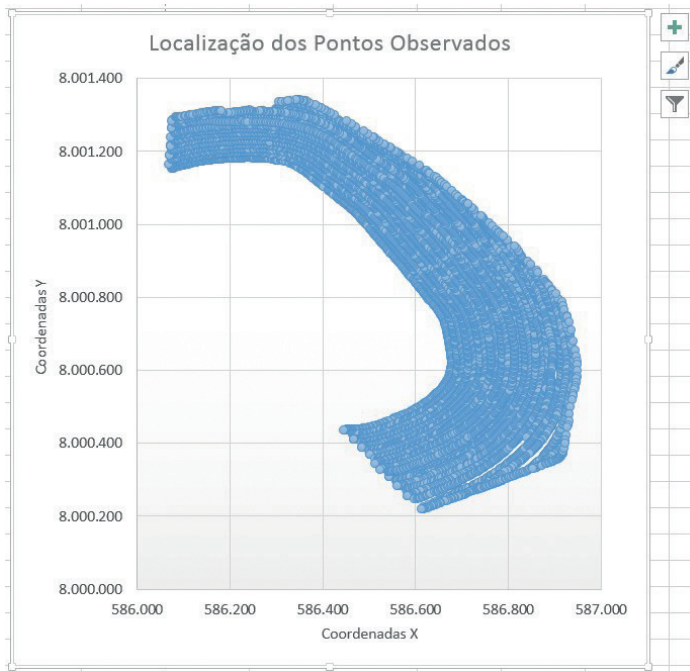
nadas retangulares em metros) para a certificação de que os pontos observados foram registrados corretamente. O exemplo no Excel utiliza os pares de coordenadas dos pontos registrados pelo GNSS, numa operação em modo contínuo das medidas de  $CE_a$  na profundidade de 0 a 30 cm. Isto permite verificar se a distribuição espacial dos pontos observados, base de toda a análise espacial, está coerente com a sua ocorrência dentro da área monitorada. O arquivo texto a ser verificado é o mesmo arquivo no formato de entrada aceito pelo Vesper, ou seja, nos formatos texto simples (“**.txt**”) ou delimitado por vírgulas (“**.csv**”).

Conforme ilustrado na Figura 2.1.a, seleciona-se as colunas com as coordenadas X e Y, nessa ordem. Na guia principal **“INSE-RIR”** da faixa de opções, clicar na seta do ícone **“Inserir Gráfico de Dispersão (X, Y) ou de Bolha”** para ver os diferentes tipos de gráficos disponíveis do grupo de comandos **“Gráficos”**. Selecionar a opção **“Dispersão”** (Figura 2.1.b). Alternativamente, o mesmo resultado pode ser obtido ao se clicar na seta **“Ver todos os Gráficos”** do grupo de comandos **“Gráficos”** (Figura 2.1.c). Na janela **“Inserir Gráfico”**, selecionar a aba **“Todos os Gráficos”**, selecionar o ícone **“Dispersão”**, nas opções de comando do menu **“XY (Dispersão)”**, e confirmar **“OK”** (Figura 2.1.c).

O resultado é um gráfico que simula um mapa onde todos os pontos do arquivo de entrada são plotados em relação às coordenadas retangulares em metros (Figura 2.2). Esta visualização gráfica simula um mapa dos pontos observados, sem ser fiel às proporções geométricas da projeção cartográfica considerada. Entretanto, a figura plotada nos permite validar a coerência dos dados de entrada em relação ao posicionamento cartográfico das coordenadas retangulares em metros, onde, pontos que venham ocorrer fora da área de estudo poderão ser desconsiderados, evitando distorções durante os subsequentes procedimentos de interpolação.



**Figura 2.1.** Opções de interface (em b e c) disponíveis na caixa de opções “Gráficos”, na aba “INserir” (em a), para gerar o gráfico de dispersão com a localização das observações em coordenadas retangulares.



**Figura 2.2.** Resultado do gráfico de dispersão ilustrando a localização dos pontos observados em coordenadas retangulares.

## 2.3 Estatística Descritiva

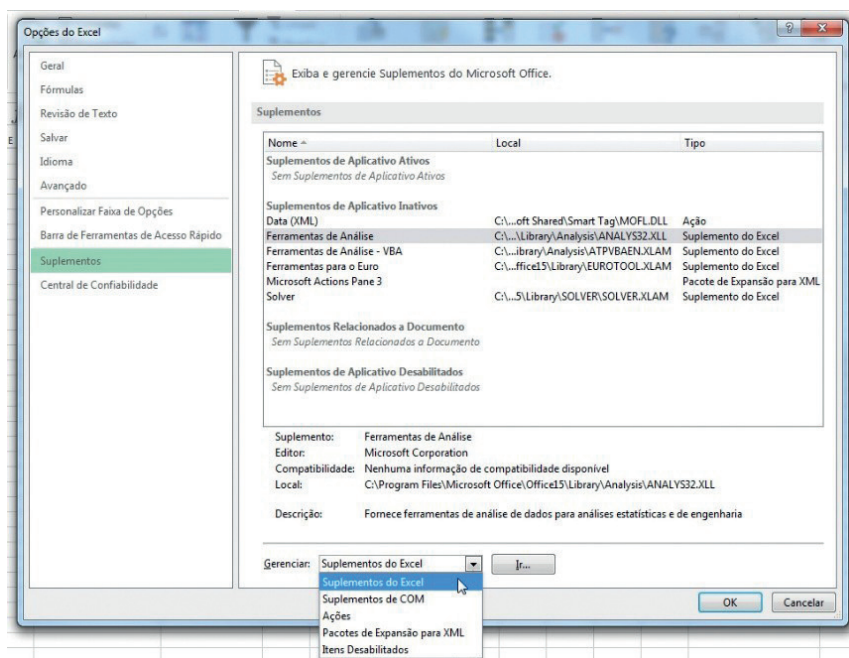
A análise exploratória dos valores absolutos tem o objetivo de identificar os dados discrepantes, a normalidade de distribuição de frequência e a variação nos dados.

Para viabilizar os procedimentos da estatística descritiva dos dados da variável  $CE_a$  no Excel, primeiro é necessário ativar as funções das ferramentas de análise de dados, que a princípio deve aparecer no grupo Suplementos de Aplicativos Inativos. Os seguintes passos permitirão habilitar o suplemento “**Ferramentas de Análise**”:

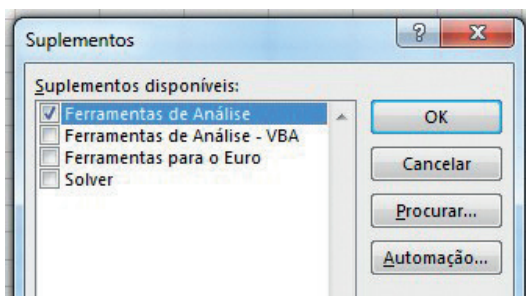
1. Na aba “**Arquivo**”, selecionar: “**Opções**” que irá abrir a janela: “**Opções do Excel**” (Figura 2.3);
2. Nesta janela clicar na opção “**Suplementos**”, que permitirá visualizar as opções de gerenciamento das funções suplementares do Excel nesta aba;

3. Ainda na aba **Suplementos**, selecionar “**Suplementos do Excel**” na barra de rolagem da opção “**Gerenciar**” e clicar no botão “**IR**” para abrir a sub-janela “**Suplementos**”;
4. Das opções listadas em “**Suplementos Disponíveis**” na sub-janela “**Suplementos**” (Figura 2.4) e selecionar “**Ferramentas de Análise**” e clicar “**OK**”.

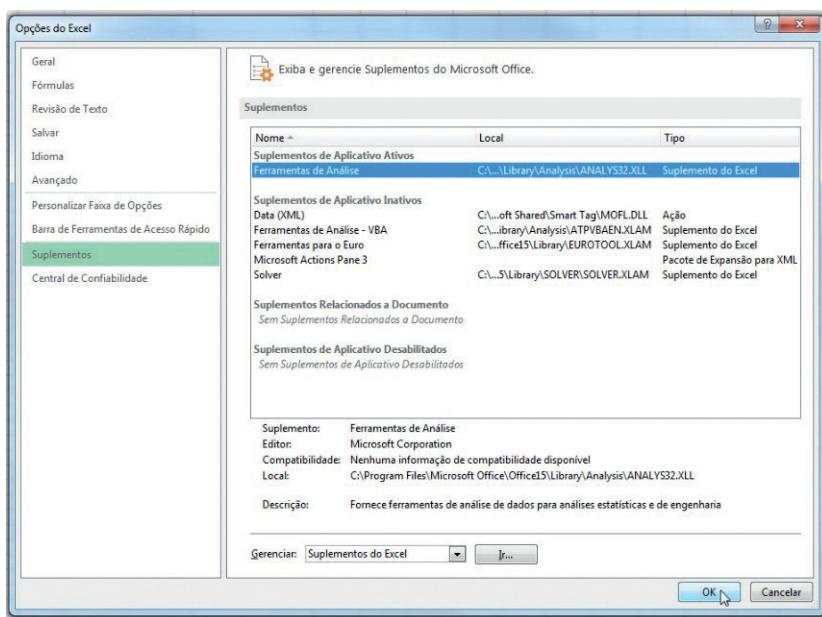
Após estes passos, o suplemento “**Ferramentas de Análise**” estará habilitado e deverá aparecer no grupo “**Suplementos de Aplicativos Ativos**” na aba “**Suplementos**” na janela “**Opções do Excel**”, condição que pode ser confirmada repetindo-se os passos 1 e 2 acima (Figura 2.5).



**Figura 2.3.** Ativação de “**Suplementos do Excel**” na janela “**Opções do Excel**” para habilitar as funções de estatística descritiva dos dados.



**Figura 2.4.** Sub-janela de “**Suplementos Disponíveis**” para ativar a opção de “**Ferramentas de Análise**” que habilitam as funções da estatística descritiva dos dados.



**Figura 2.5.** Consulta na aba “**Suplementos**” para confirmar a ativação das funções de “**Ferramentas de Análise**” listada em “**Suplementos de Aplicativos Ativos**”.

Uma vez ativado, o suplemento de análise de dados aparecerá na caixa de opções “**Análise**” da guia principal “**Dados**”. Para proceder a análise da distribuição da variável  $CE_a$ , os seguintes passos são necessários:

1. Na guia **“Dados”**, clicar na opção **“Análise de Dados”** da caixa **“Análise”** para abrir a sub-janela **“Análise de Dados”** (Figura 2.6);
2. Na barra de rolagem de **“Ferramentas de Análise”** da sub-janela **“Análise de Dados”**, selecionar a opção **“Estatística Descritiva”** e clicar em **“OK”**;
3. No campo **“Intervalo de Entrada”**, das opções de **“Entrada”**, na sub-janela **“Estatística Descritiva”**, selecionar para o cálculo a coluna, ou intervalo de células, contendo os dados da variável analisada, neste caso  $CE_a$  (Figura 2.7);
4. No campo **“Intervalo de Saída”**, das opções de **“Opções de Saída”**, na sub-janela **“Estatística Descritiva”**, selecionar o local para saída dos resultados;
5. Ainda em **“Opções de Saída”**, clicar em **“Resumo Estatístico”** e depois em **“OK”**.

Os parâmetros calculados são: média, mediana, moda, desvio padrão, variância, curtose, assimetria, valor mínimo, valor máximo e a contagem de pontos. O coeficiente de variação (CV) pode ser adicionado a este resumo de informações, sendo calculado pela Equação 1.

$$CV(\%) = \left( \frac{\text{Desvio Padrão}}{\text{Média}} \right) * 100 \quad (1)$$

Os coeficientes de assimetria e curtose expressam a normalidade de frequência dos dados e valores próximos de zero indicam distribuição de frequência normal nos dados (WEBSTER; OLIVER, 2007). A distribuição de frequência dos dados pode ser visualizada por meio do histograma que indica se uma distribuição aproxima-se de uma função normal ou não.



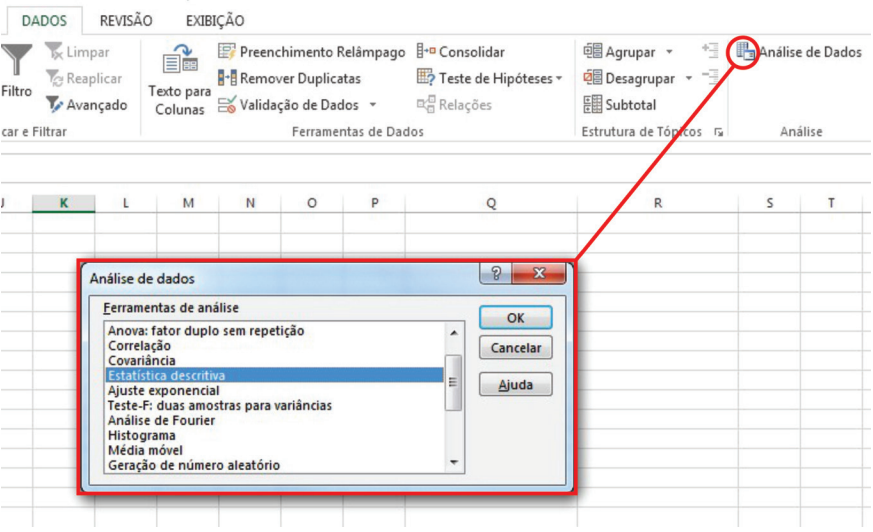


Figura 2.6. Seleção da função “Estatística Descritiva” na sub-janela de “Análise de Dados” ativada na caixa de opções “Análise” da guia principal “Dados”.

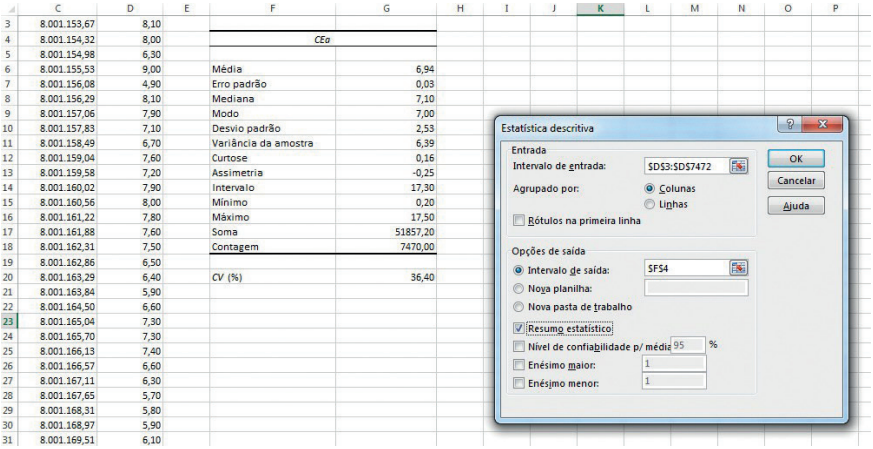


Figura 2.7. Seleção de opções de entrada e saída para o cálculo do resumo descritivo na sub-janela de “Estatística Descritiva”.

## 2.4 Geração do Histograma

O histograma é uma representação gráfica da distribuição de frequências das medições, em que o eixo das abcissas (X) corresponde ao intervalo de classe e o eixo das ordenadas (Y) à respectiva frequência.

A seguir, apresentamos um roteiro para construção do histograma no aplicativo Excel:

1. Após calcular a estatística descritiva, na guia **“Dados”**, clicar na opção **“Análise de Dados”** da caixa **“Análise”** para abrir a sub-janela **“Análise de Dados”**;
2. Na barra de rolagem de **“Ferramentas de Análise”** da sub-janela **“Análise de Dados”**, selecionar a opção **“Histograma”** e clicar em **“OK”** (Figura 2.8);
3. Com a sub-janela **“Histograma”** aberta, observar nos resultados da estatística descritiva, como os valores máximos e mínimos (Figura 2.9);
4. Criar um espaço de células e estabelecer um intervalo constante de blocos, entre o valor máximo e mínimo. No exemplo da Figura 2.9, sendo: mínimo = 0,2 e máximo = 17,5 com um intervalo de blocos variando de 2,88;
5. No campo **“Intervalo de Entrada”**, das opções de **“Entrada”**, na sub-janela **“Histograma”**, selecionar para o cálculo a coluna, ou intervalo de células, contendo os dados da variável analisada, neste caso  $CE_a$  (Figura 2.9);
6. No campo **“Intervalo do Bloco”**, das opções de **“Entrada”**, na sub-janela **“Histograma”**, selecionar para o cálculo as células contendo o intervalo de blocos (Figura 2.9) definido no passo quatro (4);
7. No campo **“Intervalo de Saída”**, das **“Opções de Saída”**, na sub-janela **“Histograma”**, indicar o local na planilha para saída dos resultados; e
8. Selecionar a opção **“Resultado no Gráfico”**, das **“Opções de Saída”**, na sub-janela **“Histograma”** e clicar em **“OK”**.

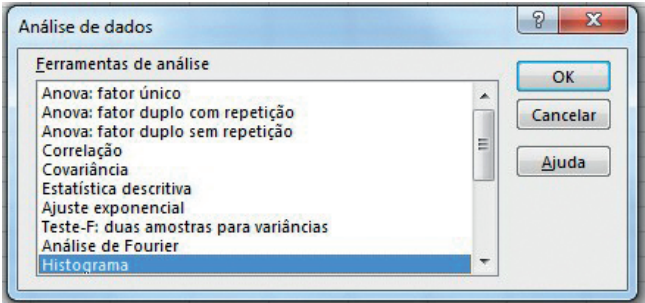


Figura 2.8. Seleção da função “Histograma” entre as opções de “Ferramentas de Análise” na sub-janela de “Análise de Dados”.

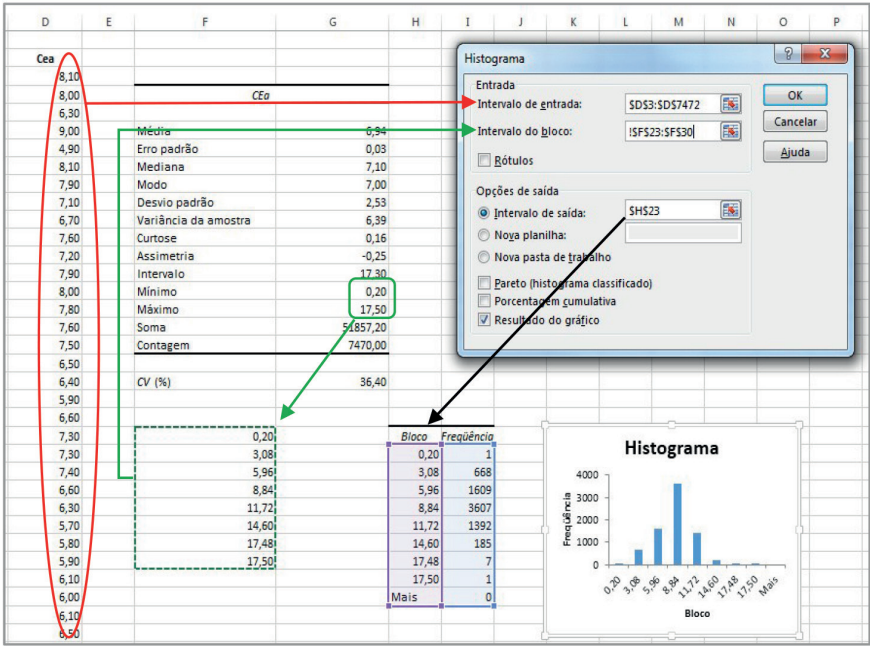


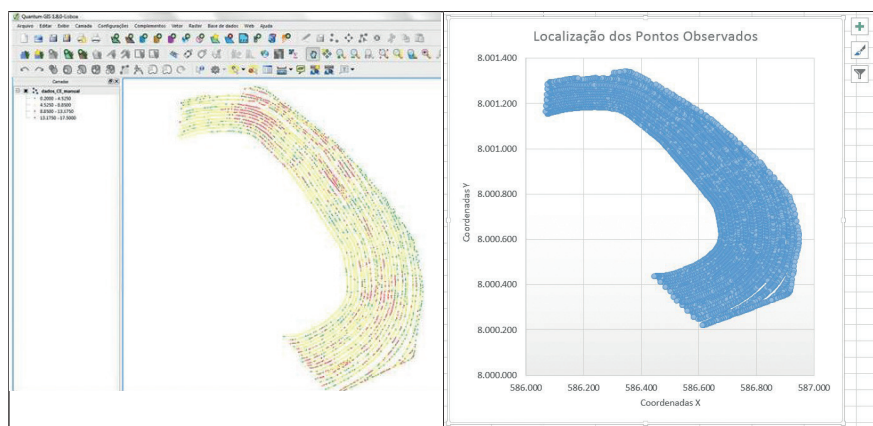
Figura 2.9. Seleção de células de “Entrada” na sub-janela de “Histograma”, para o “Intervalo de bloco” e o “Intervalo de entrada” com valores da CE<sub>a</sub>, e de células de “Saída” para o “Intervalo de saída” com o cálculo das frequências e a construção do histograma.

## 2.5 Dados Georreferenciados

Os acervos de dados georreferenciados podem ser visualizados, classificados e analisados conforme sua localização espacial e os valores dos pontos de vizinhança. Para tanto é necessário utilizar um SIG para a visualização dos dados de entrada em concordância com o sistema de coordenadas cartográficas nos quais eles foram gerados ou projetados. Neste manual optamos por utilizar como ferramental de SIG o software QGIS, o qual é apresentado em mais detalhe no sétimo capítulo. O QGIS é um SIG livre e amigável, derivado de um projeto oficial da Open Source Geospatial Foundation (OSGeo).

A visualização espacial dos dados de entrada permite obter uma primeira impressão da distribuição no espaço e identificar possíveis concentrações de valores extremos em determinado local da área de estudo. Este é um fator importante a ser considerado na análise estatística, meramente numérica, de dados discrepantes. Pois no caso de dados georreferenciados, valores extremos ou desagrupados, que fiquem fora do intervalo interquartil, podem não significar uma ocorrência errônea, sendo o seu valor aparentemente discrepante uma medida real e correta justificada pelas características da sua localização. Casos bastante comuns no monitoramento intensivo da lavoura por efeitos de borda do tráfego operacional, sombreamento de plantas, infiltrações ou concentrações localizadas de insumos e corretivos.

A precisão cartográfica só pode ser observada depois dos dados terem sido importados em um ambiente SIG. O qual apresenta funções para definição de sistemas de coordenadas e transformações entre projeções cartográficas. A informação é apresentada numa escala cartográfica e pode gerar mapas com diferentes legendas de classificação. Como ilustrado na Figura 2.10, a visualização dos dados de entrada por intermédio do gráfico de dispersão, como apresentado no início deste capítulo (i.e.: Item 2.2), é apenas uma representação de referência e serve para uma crítica espacial preliminar quanto à localização dos pontos observados, uma vez contendo distorções geométricas em sua representação. O que difere da visualização cartograficamente precisa em ambiente SIG.



(a)

(b)

**Figura 2.10.** Diferenças na visualização espacial dos pontos georreferenciados em dois ambientes de software: a) mapa de pontos com precisão cartográfica em SIG; e b) representação dos pontos num gráfico de dispersão plotado em planilha eletrônica.